QÜESTIIÓ, vol. 24, 2, p. 293-313, 2000

CONTRASTACIÓN DE HIPÓTESIS EN DISEÑOS MULTIVARIADOS SPLIT-PLOT CON MATRICES DE DISPERSIÓN ARBITRARIAS⋆

G. VALLEJO SECO
J. R. ESCUDERO GARCÍA
A. M. FIDALGO ALISTE
M. P. FERNÁNDEZ GARCÍA
Universidad de Oviedo*

El presente trabajo examina diversos procedimientos para contrastar hipótesis nulas globales, correspondientes a datos obtenidos mediante diseños multivariados split-plot cuando se incumple el supuesto de homogeneidad de las matrices de dispersión. Un examen de estos procedimientos para un amplio número de variables confirma, por un lado, la robustez del procedimiento multivariado de Welch-James dado por Johansen (1980) para probar el efecto principal de los ensayos y, por otro, la robustez de la generalización multivariada del procedimiento de Brown-Forsythe (1974) para probar la interacción de los grupos x los ensayos. Nuestros resultados también ponen de relieve que las diferencias de potencia eran pequeñas en aquellas condiciones en que tanto el procedimiento de Welch-James como de Brown-Forsythe controlaban las tasas de error de Tipo I.

Testing of hypotheses in multivariate split-plot designs with arbitrary dispersion matrices

Palabras clave: diseños multivariados split-plot, matrices de dispersión arbitrarias, procedimientos robustos y potentes

Clasificación AMS (MSC 2000): 62K10, 62J1

⋆ Este trabajo ha sido realizado con la ayuda de la DGCYT (PS95-0228).
* Departamento de Psicología. Universidad de Oviedo. Plaza Feijoo, s/n. 33003 Oviedo.
gvallejo@sci.cpd.uniovi.es.
– Recibido en marzo de 1999.
– Aceptado en marzo de 2000.

1. INTRODUCCIÓN

Los diseños de medidas repetidas constituyen un esquema de acción que se utiliza con mucha frecuencia en las investigaciones sociales y de la salud. La naturaleza de estos diseños es típicamente multivariada, sin embargo, cuando se cumplen los supuestos de normalidad conjunta multivariada, homogeneidad de las matrices de dispersión, independencia de las puntuaciones e igualdad de las varianzas correspondientes a las diferencias entre las ocasiones de medida (esfericidad), tales diseños suelen ser analizados mediante el modelo mixto de Scheffé (1956). Cuando el supuesto de esfericidad no es satisfecho, el análisis puede abordarse utilizando el enfoque univariado con los grados de libertad corregidos y/o el enfoque multivariado. De cumplirse con los supuestos exigidos por el modelo mixto, el enfoque univariado debería ser el preferido de los investigadores, pues es más potente que el enfoque multivariado (Davidson, 1972) y exige un menor número de réplicas.

Cuando los diseños de medidas repetidas contengan dos o más variables dependientes que presenten relación estadística, podemos extender los enfoques anteriores y emplear la versión multivariada de los mismos, esto es, utilizar el modelo mixto multivariado (MMM) y el modelo doblemente multivariado (MDM). Para su correcta aplicación ambos enfoques requieren que se satisfagan los supuestos de normalidad, homogeneidad e independencia, además el enfoque *MMM* requiere satisfacer el supuesto de esfericidad multivariada.

En diversos trabajos de investigación (Boik, 1991; Vallejo y Menéndez, 1997; Vallejo, Fidalgo y Fernández, 1998), se ha puesto de manifiesto que el enfoque *MMM* sin ajustar no es recomendable, a menos que el supuesto de esfericidad multivariada sea satisfecho. Cuando se cumplía dicho supuesto y el tamaño de muestra era reducido, el enfoque *MMM* ajustado mediante algún corrector tipo Box era más potente que el enfoque *MDM*. Los resultados obtenidos en estos trabajos también permiten afirmar que de los correctores tipo Box examinados, únicamente el propuesto por Boik (1991) tenía un comportamiento robusto cuando se violaba el supuesto de esfericidad, aunque tan sólo si la matriz de dispersión presentaba estructura de Kronecker. Sin embargo, conforme la matriz de dispersión se desviaba de la estructura descrita, el enfoque *MMM* ajustado mediante el corrector propuesto por Boik se volvía conservador. Por su parte, el enfoque *MDM* mantenía la tasa de error controlada al nivel nominal estipulado con independencia de la forma de la matriz de dispersión. Por consiguiente, salvo que la matriz de dispersión sea esférica y el tamaño de muestra excesivamente reducido, el enfoque *MDM* debería constituir la opción preferida por los investigadores.

Recientemente, Vallejo, Fernández, Fidalgo y Escudero (1999) evaluaron la robustez de los enfoques *MDM* y *MMM* ajustado mediante el corrector propuesto por Boik cuando las matrices de dispersión diferían entre sí y los datos carecían de normalidad. Los autores encontraron que ambos enfoques dejaban de controlar las tasas de error de Tipo

I, particularmente el enfoque *MDM*, y perdían sensibilidad estadística a medida que se incrementaba el grado de heterogeneidad de las matrices de dispersión. Si bien, cuando el diseño estaba equilibrado, la liberalidad de ambos enfoques disminuía conforme aumentaba el tamaño de muestra. El trabajo también puso de relieve el escaso efecto que la ausencia de normalidad de las puntuaciones tenía en la potencia de las pruebas y sobre el control que éstas ejercían de las tasas de error. En concreto, se encontró que el enfoque *MDM* tenía un comportamiento ligeramente liberal cuando el modelo era aditivo y conservador cuando era no aditivo; mientras que el enfoque *MMM* ajustado apenas se vio afectado.

De las investigaciones previas se desprende que cuando un investigador se encuentre inmerso en alguna situación en la cual las matrices de dispersión difieran entre sí, los datos carezcan de la normalidad requerida y los grupos estén desequilibrados, lo mejor que puede hacer es modificar el nivel de significación del enfoque *MDM* o utilizar algún procedimiento que sea robusto al incumplimiento de los supuestos del modelo, en especial, a la falta de homogeneidad de las matrices de dispersión. Por ejemplo, Keselman, Carriere y Lix (1993) sobre la base de los resultados presentados por Johansen (1980), desarrollaron una prueba multivariada Welch-James (WJ) que proporciona un buen control de la tasa de error de Tipo I para los efectos principales de un diseño univariado split-plot. Resultados similares han sido encontrados por Keselman y Lix (1997) en la perspectiva multivariada. Sin embargo, ambos estudios, cuestionan seriamente la robustez de la prueba cuando el interés se centra en los efectos de la interacción, sobre manera, cuando existe una relación inversa entre el tamaño de los grupos y el tamaño de las matrices de dispersión. En orden a solucionar los problemas reseñados, una alternativa que puede resultar prometedora, si nos atenemos a los resultados obtenidos por Algina (1994) en el contexto de los diseños univariados split-plot, es la generalización de la prueba de Brown y Forsythe (BF; 1974) al ámbito de los diseños multivariados split-plot.

En consecuencia, el objetivo fundamental del presente trabajo radica en examinar, tanto bajo hipótesis nula como bajo hipótesis alternativa, el comportamiento de los procedimientos *WJ* y *BF* cuando se contraviene el supuesto de homogeneidad de las matrices de dispersión en un diseño multivariado split-plot. De manera especial, nos interesa comprobar el funcionamiento del enfoque *BF* a la hora de contrastar los efectos de la interacción. Como ya se ha señalado, la robustez del enfoque *WJ* ha sido examinada recientemente por Keselman y Lix (1997), aunque sólo bajo hipótesis nula. Por tanto, de proporcionar ambos procedimientos un control similar de las tasas de error de Tipo I, el examen de la robustez de la potencia puede resultar decisivo. Otro objetivo del trabajo reside en investigar la posibilidad de obtener estimaciones robustas mediante el enfoque *MDM*, sabiendo cuando y cómo modificar los niveles de significación.

2. DEFINICIÓN DE LOS PROCEDIMIENTOS ESTADÍSTICOS

Considérese una situación en la cual N unidades muestrales independientes estén repartidas en p grupos, posiblemente no equilibrados, $n_j,\ j = 1, \dots, p$, y $\sum_{j=1}^{p} n_j = N$. Asumamos también que para cada unidad, se registren r características diferentes bajo cada una de las q ocasiones de medida. El resultado para cada una de las unidades que configuran el diseño es una respuesta r-dimensional en q ocasiones que puede ser descrita dentro del contexto del modelo lineal general como sigue:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{U}, \tag{1}$$

donde $\boldsymbol{Y}$ es la matriz de respuestas de orden $N \times t$, $t = qr$, $\boldsymbol{X}$ es la matriz de diseño de orden $N \times p$ con rango $(\boldsymbol{X}) = p$, $\boldsymbol{B}$ es la matriz de parámetros no aleatorios (p.e., medias de la población) de orden $p \times t$ y $\boldsymbol{U}$ es la matriz de errores de orden $N \times t$. Las respuestas son ordenadas por columnas conforme a las variables dependientes y dentro de éstas conforme a las ocasiones de medida. Si denotamos por $\boldsymbol{u}_i'$ el vector de errores aleatorios correspondiente al sujeto i-ésimo, es asumido que

$$\boldsymbol{u}_i' \sim N(\boldsymbol{0}, \Sigma), \tag{2}$$

donde la matriz de covarianzas Σ es una matriz definida positiva de orden $t \times t$. El hecho de que la forma Σ no dependa de i supone que todos los vectores de errores aleatorios $\boldsymbol{u}$ tienen la misma matriz Σ y, por ende, que dichas matrices son homogéneas. Sin embargo, puede ocurrir que no todos los vectores de errores tengan la misma matriz Σ, en estos casos, puede resultar más adecuado usar pruebas que no requieran satisfacer el supuesto de igualdad de las matrices de dispersión. Por tal motivo, en los apartados que siguen se describen tres enfoques diferentes. El primero es pertinente cuando se cumple que la forma de Σ no depende de la población de la que se extraen las n_j unidades experimentales, mientras que los dos restantes lo son cuando el supuesto no se satisface.

2.1. Enfoque del modelo doblemente multivariado

Para probar que h funciones estimables tienen un valor especificado, la hipótesis nula, H_0, es expresada como sigue:

$$H_0 : \boldsymbol{C}'\boldsymbol{B}\boldsymbol{A} = \boldsymbol{0}, \tag{3}$$

donde la matriz $\boldsymbol{C}'$, de dimensión $h \times p$ con rango $(\boldsymbol{C}) = h$, es usada para definir un conjunto de h contrastes entre los grupos de tratamiento y la matriz $\boldsymbol{A}$, de dimensión $t \times k$ con rango $(\boldsymbol{A}) = k$, es usada para definir un conjunto de k contrastes entre las diferentes ocasiones de medida para cada una de las r variables dependientes. La matriz $\boldsymbol{B}$ ya ha sido definida con anterioridad.

Los estadísticos usados para comprobar las hipótesis de interés son función de las raíces características de $\boldsymbol{HE}^{-1}$, donde las matrices sumas de cuadrados y productos cruzados (*SCPC*) correspondientes a la hipótesis y al error son obtenidas como sigue:

$$\text{(4)} \qquad \boldsymbol{H} = (\boldsymbol{C}'\hat{\boldsymbol{B}}\boldsymbol{A})'[\boldsymbol{C}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}]^{-1}(\boldsymbol{C}'\hat{\boldsymbol{B}}\boldsymbol{A}) \quad \text{y} \quad \boldsymbol{E} = \boldsymbol{A}'\boldsymbol{Y}'[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{Y}\boldsymbol{A}.$$

Denotando por $W_t = (\nu, \boldsymbol{A}'\Sigma\boldsymbol{A}, \Xi)$ la distribución Wishart no central de dimensión t con ν grados de libertad, matriz de covarianza $\boldsymbol{A}'\Sigma\boldsymbol{A}$ y matriz de no centralidad $\Xi[\Xi = (\boldsymbol{A}'\Sigma\boldsymbol{A})^{-1}\Phi]$, Boik (1988) muestra que $\boldsymbol{H}$ y $\boldsymbol{E}$ se distribuyen independientemente como sigue:

$$\text{(5)} \qquad \boldsymbol{H} \sim W_t[\nu_h, \boldsymbol{A}'\Sigma\boldsymbol{A}, (\boldsymbol{A}'\Sigma\boldsymbol{A})^{-1}\Phi] \quad \text{y} \quad \boldsymbol{E} \sim W_t[\nu_e, \boldsymbol{A}'\Sigma\boldsymbol{A}, \boldsymbol{0}],$$

donde $\nu_h = h =$ rango de $\boldsymbol{C}$ y $\nu_e = N - p$ son, respectivamente, los grados de libertad de matrices $\boldsymbol{H}$ y $\boldsymbol{E}$, $\Phi = (\boldsymbol{C}'\boldsymbol{B}\boldsymbol{A})'[\boldsymbol{C}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}]^{-1}(\boldsymbol{C}'\boldsymbol{B}\boldsymbol{A})$ y $\Sigma = (N-p)^{-1}\{\boldsymbol{Y}'[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{Y}\}$.

Habiendo computado las matrices *SCPC* correspondientes a la hipótesis $\boldsymbol{H}$ y al error $\boldsymbol{E}$ cualquiera de los criterios estadísticos existentes (p.e, la traza generalizada de Hotelling o lambda de Wilks) puede ser utilizado para comprobar la veracidad de la hipótesis nula.

La hipótesis de ausencia de interacción entre los grupos y las ocasiones de medida del diseño multivariado split-plot se prueba definiendo las matrices de contrastes $\boldsymbol{C}'$ y $\boldsymbol{A}$ como

$$\text{(6)} \qquad \underset{(p-1)\times p}{\boldsymbol{C}'} = [\boldsymbol{I}_{p-1} \vdots -\boldsymbol{1}] \quad \text{y} \quad \underset{t\times(q-1)r}{\boldsymbol{A}} = \boldsymbol{I}_r \otimes \begin{bmatrix} \boldsymbol{I}_{q-1} \\ \vdots \\ -\boldsymbol{1} \end{bmatrix},$$

donde $\otimes$ denota el producto de Kronecker. A su vez, para probar la hipótesis nula multivariada de igualdad de las ocasiones de medida (modelo noaditivo y medias no ponderadas), la matriz $\boldsymbol{A}$ se define como en la ecuación (6) y $\boldsymbol{C}$ es una matriz de identidad de orden $p \times p$.

Ambas hipótesis fueron probadas usando la aproximación F de Rao (1951) a la lambda de Wilks (1932) que sigue:

$$\text{(7)} \qquad F = \frac{1 - \Lambda^{1/s}}{\Lambda^{1/s}} \left(\frac{\nu_2}{\nu_1} \right),$$

donde $s = [(m^2\nu_h^2 - 4)/(m^2 + \nu_h^2 - 5)]^{1/2}$, $\nu_1 = m\nu_h, \nu_2 = \{[\nu_e - (m - \nu_h + 1)/2]s - (m\nu_h - 2)/2\}$, y $\Lambda = |\boldsymbol{E}|/|\boldsymbol{E} + \boldsymbol{H}|$, con m igual a la dimensión de $\boldsymbol{E}$ y $\boldsymbol{H}$. Cada una de estas hipótesis fue rechazada al nivel α si $F > F_{(1-\alpha);\nu_1,\nu_2}$, donde $F_{(1-\alpha);\nu_1,\nu_2}$ es el $100(1-\alpha)$ ésimo percentil de la distribución F con ν_1 y ν_2 y grados de libertad.

2.2. El procedimiento de Brown-Forsythe

Para probar la hipótesis de igualdad de p medias cuando las varianzas de la población son heterogéneas, Brown y Forsythe (1974) propusieron el estadístico

$$F^{*} = \frac{(\boldsymbol{C}'\hat{\beta})'[\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}'(\boldsymbol{C}'\hat{\beta})/(p-1)}{\left[\sum_{j=1}^{p} c_j \hat{\sigma}_j^2\right]}, \tag{8}$$

donde $\hat{\sigma}_j^2 = \boldsymbol{y}_j'[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{y}_j$ y $c_j = (1 - n_j/N)$. El estadístico F^* se distribuye aproximadamente como F con $p-1$ y f^* grados de libertad, donde

$$f^{*} = \left[\sum_{j=1}^{p} c_j \sigma_j^2\right]^2 \cdot \left[\sum_{j=1}^{p} \frac{c_j^2 \sigma_j^4}{(n_j - 1)}\right]^{-1} \tag{9}$$

es determinado usando el método de Satterthwaite (1941).

Coombs y Algina (1992, 1996) han extendido el enfoque de Brown y Forsythe al ámbito multivariado, reemplazando medias y varianzas por las matrices *SCPC* correspondientes a la hipótesis y al error. Para el diseño propuesto, los estadísticos usados para contrastar las hipótesis de interés son función de las raíces características de $\boldsymbol{H}\tilde{\boldsymbol{E}}^{-1}$. Donde

$$\boldsymbol{H} = (\boldsymbol{C}'\hat{\boldsymbol{B}}\boldsymbol{A})'[\boldsymbol{C}'\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{C}]^{-1}(\boldsymbol{C}'\hat{\boldsymbol{B}}\boldsymbol{A}) \quad \text{y} \quad \tilde{\boldsymbol{E}} = \left(\frac{\nu_e^*}{p-1}\right)\sum_{j=1}^{p} c_j \boldsymbol{A}'\Sigma_j \boldsymbol{A}. \tag{10}$$

Aplicando los resultados de Nel y van der Merwe (1986), la matriz $\sum_{j=1}^{p} c_j \boldsymbol{A}'\Sigma_j \boldsymbol{A}$ puede ser aproximada como una suma de distribuciones Wishart

$$\sum_{j=1}^{p} \left(c_j \boldsymbol{A}'\Sigma_j \boldsymbol{A}\right) \sim SW_{qr}\left(\nu_1^*, \ldots, \nu_p^*; \frac{c_1}{\nu_1^*}\boldsymbol{A}'\Sigma_1\boldsymbol{A}, \ldots, \frac{c_p}{\nu_p^*}\boldsymbol{A}'\Sigma_p\boldsymbol{A}\right), \tag{11}$$

con grados de libertad

$$\nu_e^* = \frac{tr\left[\sum_{j=1}^{p}\left(c_j \boldsymbol{A}'\Sigma_j \boldsymbol{A}\right)\right]^2 + \left[tr\sum_{j=1}^{p}\left(c_j \boldsymbol{A}'\Sigma_j \boldsymbol{A}\right)\right]^2}{\sum_{j=1}^{p}\frac{1}{n_j - 1}\left\{tr\left(c_j \boldsymbol{A}'\Sigma_j \boldsymbol{A}\right)^2 + \left[tr\left(c_j \boldsymbol{A}'\Sigma_j \boldsymbol{A}\right)\right]^2\right\}}, \tag{12}$$

donde *tr* denota el operador traza.

Las dos hipótesis de interés del diseño se prueban definiendo las matrices $\boldsymbol{C}'$ y $\boldsymbol{A}$ en los mismos términos que bajo el enfoque *MDM* y utilizando, con las modificaciones oportunas, la aproximación F a la lambda de Wilk propuesta por Rao que sigue:

$$F = \frac{1-\Lambda^{1/s}}{\Lambda^{1/s}}\left(\frac{\nu_2^*}{\nu_1}\right), \tag{13}$$

donde $s = [(m^2\nu_h^2 - 4)/(m^2 + \nu_h^2 - 5)]^{1/2}, \nu_1 = m\nu_h$ y $\nu_2^* = \{[\nu_e^* - (m - \nu_h + 1)/2]s - (m\nu_h - 2)/2\}$, con m igual a la dimensión de $\tilde{\boldsymbol{E}}$ y $\boldsymbol{H}$. Cada una de estas hipótesis se rechaza al nivel α si $F > F_{(1-\alpha);\nu_1,\nu_2^*}$, donde $F_{(1-\alpha);\nu_1,\nu_2^*}$ es el $100(1-\alpha)$ ésimo percentil de la distribución F con ν_1 y ν_2^* grados de libertad.

2.3. El procedimiento de Welch-James

La otra solución propuesta para probar la igualdad de vectores de medias cuando las matrices de dispersión difieren entre sí, es la versión multivariada del enfoque Welch-James desarrollada por Keselman *et al.* (1993) sobre la base del trabajo de Johansen (1980). La hipótesis lineal general para esta solución es

$$H_0 :\ \boldsymbol{R}\mu = \boldsymbol{0} \tag{14}$$

donde $\boldsymbol{R} = [\boldsymbol{C}' \otimes (\boldsymbol{I}_r \otimes \boldsymbol{A}')]$ es una matriz de contrastes cuyo orden depende de la hipótesis que estemos probando y $\mu = \text{vec}(\mu') = (\mu_1 \ldots \mu_p)'$, con $\mu_j = [\mu_{j1} \ldots \mu_{jt}]$. Así pues, μ es un vector columna con los pt elementos obtenidos tras concatenar verticalmente los vectores de medias μ_j. La prueba estadística viene dada por

$$T_{W-J} = (\boldsymbol{R}\hat{\mu})'(\boldsymbol{R}\boldsymbol{P}\boldsymbol{R}')^{-1}(\boldsymbol{R}\hat{\mu}), \tag{15}$$

donde $\boldsymbol{P}$ es una matriz diagonal a bloques de orden $pt \times pt$ con cada una de las matrices de dispersión muestrales ponderadas por n_j a lo largo de la diagonal principal. El estadístico T_{WJ} dividido por una constante, c, es convenientemente aproximado mediante la distribución F con grados de libertad ν_1 (rango de la matriz $\boldsymbol{R}$) y $\nu_2 = \nu_1(\nu_1 + 2)/3A$. La constante $c = \nu_1 + 2A - 6A/(\nu_1 + 2)$, con

$$A = \frac{1}{2}\sum_{j=1}^{p}\left[\text{tr}\left\{\boldsymbol{P}\boldsymbol{R}'(\boldsymbol{R}\boldsymbol{P}\boldsymbol{R}')^{-1}\boldsymbol{R}\boldsymbol{Q}_j\right\}^2 + \left\{\text{tr}(\boldsymbol{P}\boldsymbol{R}'(\boldsymbol{R}\boldsymbol{P}\boldsymbol{R}')^{-1}\boldsymbol{R}\boldsymbol{Q}_j)\right\}^2\right]/(n_j - 1), \tag{16}$$

donde $\boldsymbol{Q}_j$ es una matriz diagonal de bloques de orden $pt \times pt$ con el j-ésimo bloque igual a una matriz de identidad de orden $t \times t$ y el resto ceros.

Como muestran Vallejo y Escudero (1998), la forma de $\boldsymbol{R}$ para probar la hipótesis lineal general $H_0 :\ \boldsymbol{R}\mu = \boldsymbol{0}$ depende del efecto que se desee contrastar. Para el caso concreto de la interacción, $\boldsymbol{R} = [\boldsymbol{C}' \otimes (\boldsymbol{I}_r \otimes \boldsymbol{A}')]$ es una matriz de orden $(p-1)(q-1)r \times pt$, donde

$\boldsymbol{C}'$ es una matriz de coeficientes de orden $(p-1)\times p$ que determina los elementos de $\boldsymbol{B}$ a incluir en H_0, $\boldsymbol{A}$ es una matriz de coeficientes de orden $q\times(q-1)$ que nos permite generar hipótesis entre los diferentes niveles de la variable ocasiones de medida e $\boldsymbol{I}_r$ es una matriz de identidad de orden $r\times r$. Mientras que para la hipótesis multivariada de igualdad de las ocasiones de medida (modelo no aditivo y medias no ponderadas), $\boldsymbol{R}=[\boldsymbol{C}'\otimes(\boldsymbol{I}_r\otimes\boldsymbol{A}')]$ es una matriz de orden $p(q-1)r\times pt$, donde $\boldsymbol{C}'$ es una matriz de identidad de orden $p\times p$, $\boldsymbol{A}$ es una matriz de contrastes de orden $q\times(q-1)$ e $\boldsymbol{I}_r$ es una matriz de identidad de orden $r\times r$.

Para ambos efectos la $H_0:\boldsymbol{R}\mu=\boldsymbol{0}$ se rechaza al nivel α si $T_{WJ}/c>F_{(1-\alpha);\nu_1,\nu_2}$, donde $F_{(1-\alpha);\nu_1,\nu_2}$ es el $100(1-\alpha)$ ésimo percentil de la distribución F con grados de libertad ν_1 y ν_2.

3. MÉTODO

Para investigar el desempeño de las diferentes estrategias analíticas realizamos un estudio de simulación Monte Carlo con datos muestreados desde distribuciones normales multivariadas en el que estimamos las tasas de error de Tipo I y la potencia estadística de las pruebas (bajo hipótesis alternativa el patrón de configuración de las medias fue el de rango mínimo). Para ello manipulamos el tamaño de muestra, el equilibrio de los grupos, la relación entre el tamaño de los grupos y el de las matrices de dispersión y el tipo de matrices de dispersión. Con respecto al primer criterio, las hipótesis a comparar son las referidas a diseños de medidas parcialmente repetidas (tres grupos de tratamiento, $p=3$, tres medidas repetidas, $q=3$ y tres variables dependientes, $r=3$) con vectores de observaciones del mismo tamaño ($n_1=n_2=n_3=12$ y $n_1=n_2=n_3=18$) y vectores de observaciones de tamaño distinto. Estos valores muestrales no se fijaron arbitrariamente, sino que a la vista de los resultados encontrados por Tang y Algina (1993), se seleccionaron los valores de N de modo que la razón entre éstos y el número de variables dependientes se encontrase comprendida entre diez y veinte. Además, también se tuvo en cuenta que dichos valores muestrales fueran representativos de los hallados con relativa frecuencia en las investigaciones psicológicas y educacionales.

En lo que se refiere al segundo criterio, se manipularon condiciones con iguales tamaños de muestra ($n_1=n_2=n_3=12$ y $n_1=n_2=n_3=18$) y con distinto tamaño de muestra ($n_1=9, n_2=12, n_3=15$ y $n_1=12, n_2=18, n_3=24$). Para cada valor de N manipulamos la desigualdad del tamaño de los grupos de una manera leve y moderada, en concreto, los coeficientes de variación de los tamaños muestrales oscilaron en torno a 0.20 y 0.28. A su vez, las razones entre los grupos con menor tamaño de muestra ($n_{mín}$) y el número de variables dependientes multiplicado por el de medidas repetidas menos uno se fijaron en 1.5 y 2. Dichos valores resultan habituales en los trabajos de experimentación de muchos investigadores.

Por lo que respecta al tercer criterio, investigamos el comportamiento de los enfoques cuando la naturaleza de la relación entre el tamaño de los grupos y el tamaño de las matrices de dispersión era, tanto positiva como negativa. Una relación positiva implica que el grupo de menor tamaño se asocia con la matriz de dispersión más pequeña, mientras que una relación negativa implica que el grupo de menor tamaño se asocia con la matriz de dispersión más grande. El grado de heterogeneidad de las matrices de dispersión incluido en el presente trabajo fue $\Sigma_3 = 5\Sigma_1$ y $\Sigma_2 = 3\Sigma_1$.

A su vez, por lo que respecta al cuarto criterio, las matrices de dispersión $\boldsymbol{A}'\Sigma\boldsymbol{A}$ fueron seleccionadas sobre la base de las características que siguen:

A: Estructura de Kronecker $[\boldsymbol{A}'\Sigma\boldsymbol{A} = \Psi \otimes \boldsymbol{I}]$ y el valor de ε_1 = 1.00 y el de ε_2 = 1.00
B: Estructura de Kronecker $[\boldsymbol{A}'\Sigma\boldsymbol{A} = \Psi \otimes \boldsymbol{I}]$ y el valor de ε_1 = 0.50 y el de ε_2 = 1.00
C: Estructura de Kronecker $[\boldsymbol{A}'\Sigma\boldsymbol{A} = \Psi \otimes \Omega]$ y el valor de ε_1 = 0.50 y el de ε_2 = 0.75
D: Estructura de Kronecker $[\boldsymbol{A}'\Sigma\boldsymbol{A} \neq \Psi \otimes \Omega]$ y el valor de ε_1 = 0.50 y el de ε_2 = 0.75
E: Estructura de Kronecker $[\boldsymbol{A}'\Sigma\boldsymbol{A} \neq \Psi \otimes \Omega]$ y el valor de ε_1 = 0.33 y el de ε_2 = 0.57

Las tres primeras matrices tienen estructura de Kronecker, además las matrices correspondientes a los casos *A* y *B* presentan esfericidad multivariada, mientras que la matriz correspondiente al caso *C* tiene estructura de Kronecker, pero carece de esfericidad. Las dos últimas fueron seleccionadas de manera tal que el grado de desviación de la estructura de Kronecker (δ) fuese leve, como sucede con la matriz correspondiente al caso $D(\delta = 0.014)$, o bien severo, como ocurre con la matriz correspondiente al caso $E(\delta = 0.678)$. Donde ε_1 se refiere al factor de corrección de Greenhouse y Geisser (1959) y ε_2 al factor de corrección de Boik (1991). Las matrices de dispersión con los valores especificados de epsilón se lograron mediante el algoritmo desarrollado por Cornell, Young y Bratcher (1991).

Las simulaciones fueron ejecutadas generando observaciones desde distribuciones normales. Para ello, en un principio, vectores de observaciones pseudoaleatorios $\boldsymbol{y}'_{ij1}, \boldsymbol{y}'_{ij2}, \ldots, \boldsymbol{y}'_{ijr}$, con vector de medias $\mu'_j = [\mu_{j11}, \mu_{j21}, \ldots, \mu_{jk1}, \ldots, \mu_{jkr}]$ y matriz de varianzas-covarianzas Σ, fueron obtenidos desde distribuciones normales con $\mu = 0$ y $\sigma^2 = 1$ utilizando el programa *GAUSS* (v.3.2.32). A continuación, las pertinentes observaciones multivariadas se consiguieron siguiendo el método de Schauer y Stoller (1966); esto es,

$$\boldsymbol{y}'_{ij} = \boldsymbol{T}\boldsymbol{z}_{ij} + \mu_{ij} \tag{17}$$

donde $\boldsymbol{T}$ es la factorización Cholesky de Σ_j y $\boldsymbol{z}_{ij}$ es un vector de variables normales generado de acuerdo al algoritmo propuesto por Kinderman y Ramage (1976).

Por último, mediante un programa escrito en *GAUSS* (1997) se efectuó el análisis del conjunto de datos simulados con cada uno de los enfoques expuestos; esta operación permite comparar el comportamiento de las técnicas en relación con cada una de las variables manipuladas. Las comparaciones fueron efectuadas examinando las tasas de error de Tipo I y la potencia de las pruebas. Bajo H_0 la proporción empírica de errores de Tipo I se obtuvo dividiendo el número de veces que cada estadístico excedía su valor crítico por el de réplicas efectuadas. A su vez, la potencia de prueba correspondiente a las ocasiones de medida e interacción grupos $\times$ ocasiones se obtuvo dividiendo el número de veces que la H_0 era correctamente rechazada por el número de réplicas practicadas. Para cada condición manipulada se han efectuado 10.000 replicaciones adoptando los niveles de significación α =0.01 y α =0.05.

4. RESULTADOS

Las tasas empíricas de error de Tipo I consideradas aceptables en el presente trabajo han sido aquellas que se encuentran comprendidas dentro del intervalo más/menos dos errores estándar (SE). Los errores estándar mostrados para las estimaciones empíricas de los niveles de significación se obtuvieron mediante la expresión $[\alpha(1-\alpha)/g]^{1/2}$, donde g denota el número de réplicas efectuadas. Las estimaciones empíricas que en la Tablas 1 y 2 aparecen resaltadas en negrita, son aquellas que han sido declaradas significativas por encontrase fuera de las bandas de $\pm$2SE y, por ende, representan situaciones en las que los diferentes enfoques dejan de comportarse de una manera robusta. Con 10.000 replicaciones para α =0.05, el intervalo criterio fue $(0.0456 \leqslant \hat{\alpha} \leqslant 0.0543)$, mientras que para α =0.01 el intervalo criterio fue $(0.00801 \leqslant \hat{\alpha} \leqslant 0.00119)$. A su vez, para evaluar la robustez de la potencia, se compararon los distintos procedimientos entre sí y se comprobó cual resultaba más potente cuando el supuesto de homogeneidad de las matrices no se satisfacía.

4.1. Tasas empíricas de error de Tipo I y de potencia para las ocasiones de medida

La Tabla 1 contiene las tasas empíricas de error y de potencia correspondientes a las ocasiones de medida para los tres procedimientos analíticos de interés, bajo cada una de las diferentes condiciones manipuladas. Por lo que se refiere a las tasas de error de Tipo I asociadas con el enfoque *MDM*, los datos presentados en la Tabla 1 ponen de relieve los aspectos siguientes: En primer lugar, cuando el diseño estaba equilibrado el enfoque *MDM* siempre se comportaba de un modo ligeramente liberal; por ejemplo, cuando $n_1 = n_2 = n_3 = 12$, las tasas de error empíricas promediadas a lo largo de las diferentes estructuras de covarianza fueron $\hat{\alpha} = 0.071$ y $\hat{\alpha} = 0.0159$, para cada uno de los niveles de significación nominal utilizados. Si se mantenía el equilibrio entre los grupos, la liberalidad disminuía conforme se incrementaba el tamaño de muestra. En segundo lugar, cuando el diseño estaba desequilibrado y existía una relación positiva entre el tamaño de los grupos y el de las matrices de dispersión, el enfoque *MDM* siempre se volvía conservador; por ejemplo, cuando $n_1 = 9, n_2 = 12$ y $n_3 = 15$, las tasas de error

Tabla 1. Tasas empíricas de error y de potencia para las ocasiones de medida bajo el modelo no aditivo.

				MDM		*BF*		*WJ*		*MDM*		*BF*		*WJ*	
Σ	n_1	n_2	n_3	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$
	12	12	12	0.0708	0.0147	0.0471	0.0087	0.0518	0.0096	0.5448	0.2940	0.4656	0.2103	0.5055	0.2386
	09	12	15	0.0282	0.0049	0.0424	0.0068	0.0543	0.0118	0.4083	0.1715	0.4907	0.2347	0.5588	0.2885
A	15	12	09	0.1453	0.0443	0.0426	0.0071	0.0535	0.0090	0.6432	0.3978	0.3670	0.1286	0.3983	0.1454
	18	18	18	0.0659	0.0153	0.0507	0.0100	0.0534	0.0098	0.7787	0.5373	0.7404	0.4768	0.7471	0.4720
	12	18	24	0.0128	0.0020	0.0342	0.0061	0.0469	0.0085	0.6014	0.3212	0.7477	0.4789	0.8002	0.5522
	24	18	12	0.1684	0.0589	0.0543	0.0091	0.0505	0.0082	0.8425	0.6580	0.6043	0.3180	0.5762	0.2730
	12	12	12	0.0704	0.0160	0.0475	0.0093	0.0542	0.0113	0.6451	0.3828	0.5695	0.2805	0.5966	0.3109
	09	12	15	0.0267	0.0052	0.0399	0.0073	0.0527	0.0120	0.5156	0.2486	0.6021	0.3196	0.6691	0.3871
B	15	12	09	0.1402	0.0455	0.0442	0.0648	0.0543	0.0089	0.7179	0.4808	0.4492	0.1731	0.4774	0.1972
	18	18	18	0.0627	0.0158	0.0488	0.0104	0.0501	0.0094	0.8758	0.6576	0.8256	0.5933	0.8359	0.6022
	12	18	24	0.0135	0.0024	0.0351	0.0054	0.0462	0.0089	0.7236	0.4437	0.8460	0.6271	0.8861	0.6933
	24	18	12	0.1777	0.0649	0.0538	0.0099	0.0482	0.0081	0.8998	0.7611	0.7166	0.4242	0.6753	0.3814
	12	12	12	0.0693	0.0166	0.0465	0.0089	0.0522	0.0106	0.5424	0.2909	0.4690	0.2045	0.5086	0.2416
	09	12	15	0.0284	0.0043	0.0398	0.0073	0.0549	0.0107	0.4022	0.1828	0.4877	0.2349	0.5561	0.2882
C	15	12	09	0.1407	0.0429	0.0407	0.0061	0.0527	0.0082	0.6432	0.3990	0.3712	0.1313	0.3995	0.1495
	18	18	18	0.0625	0.0147	0.0482	0.0101	0.0512	0.0098	0.7740	0.5332	0.7413	0.4774	0.7478	0.4740
	12	18	24	0.0143	0.0018	0.0355	0.0070	0.0495	0.0101	0.5955	0.3177	0.7491	0.4835	0.7991	0.5649
	24	18	12	0.1712	0.0561	0.0543	0.0111	0.0513	0.0093	0.8444	0.6563	0.6154	0.3159	0.5723	0.2814
	12	12	12	0.0717	0.0155	0.0466	0.0092	0.0547	0.0099	0.5519	0.3144	0.4737	0.2125	0.5002	0.2303
	09	12	15	0.0241	0.0046	0.0373	0.0063	0.0549	0.0103	0.4053	0.1642	0.4886	0.2228	0.5597	0.2786
D	15	12	09	0.1438	0.0482	0.0448	0.0067	0.0539	0.0094	0.6445	0.4011	0.3693	0.1312	0.4054	0.1495
	18	18	18	0.0661	0.0156	0.0507	0.0098	0.0489	0.0103	0.7728	0.5442	0.7413	0.4774	0.7363	0.4671
	12	18	24	0.0147	0.0020	0.0378	0.0069	0.0518	0.0084	0.6002	0.3140	0.7426	0.4799	0.8042	0.5623
	24	18	12	0.1698	0.0605	0.0536	0.0105	0.0498	0.0083	0.8369	0.6608	0.6052	0.3167	0.5830	0.2804
	12	12	12	0.0728	0.0171	0.0496	0.0094	0.0539	0.0104	0.5239	0.2750	0.4451	0.1985	0.4817	0.2180
	09	12	15	0.0270	0.0049	0.0398	0.0071	0.0555	0.0113	0.3723	0.1533	0.4559	0.2013	0.5272	0.2539
E	15	12	09	0.1409	0.0440	0.0432	0.0067	0.0503	0.0089	0.6207	0.3748	0.3478	0.1158	0.3777	0.1368
	18	18	18	0.0652	0.0142	0.0496	0.0102	0.0517	0.0100	0.7424	0.5039	0.7039	0.4355	0.7097	0.4358
	12	18	24	0.0149	0.0023	0.0345	0.0065	0.0508	0.0095	0.5578	0.2817	0.7314	0.4407	0.7742	0.5228
	24	18	12	0.1729	0.0593	0.0544	0.0097	0.0506	0.0084	0.8151	0.6253	0.5743	0.2912	0.5523	0.2619

MDM = Modelo doblemente multivariado; *BF* = Procedimiento de Brown-Forsythe; *WJ* = Procedimiento de Welch-James.

empíricas promediadas a través de las cinco matrices de dispersión fueron $\hat{\alpha} = 0.0268$ y $\hat{\alpha} = 0.0048$, para cada uno de los niveles de significación utilizados. El grado de conservadurismo se incrementaba conforme lo hacía el coeficiente de variación muestral. En tercer lugar, cuando el tamaño de los grupos era desigual y existía una relación negativa entre el tamaño de los grupos y el tamaño de las matrices de dispersión, el enfoque *MDM* siempre se volvía liberal; por ejemplo, cuando $n_1 = 15, n_2 = 12$ y $n_3 = 9$, las tasas de error empíricas promediadas a lo largo de las distintas matrices de dispersión fueron $\hat{\alpha} = 0.1421$ y $\hat{\alpha} = 0.0449$, para cada uno de los niveles de significación utilizados. En esta situación, a diferencia de lo que sucedía cuando el diseño estaba equilibrado, el enfoque se tornaba más liberal conforme se incrementaba el valor de N.

Con respecto a las tasas de error de Tipo I asociadas con el procedimiento *BF*, los datos de la Tabla 1 muestran dos cosas: por un lado, que cuando el diseño estaba equilibrado el procedimiento mantenía siempre las tasas de error controladas dentro del intervalo definido y, por otro lado, que el procedimiento se volvía ligeramente conservador para el resto de las condiciones investigadas. Si bien el grado de conservadurismo no permanecía estable, pues cuando la relación entre el tamaño de los grupos y el de las matrices de dispersión era positiva el grado de conservadurismo se incrementaba conforme lo hacía el coeficiente de variación muestral, mientras que cuando la relación era negativa el valor empírico de (se aproximaba estrechamente a su valor nominal al incrementarse el coeficiente de variación muestral.

A su vez, por lo que respecta a las tasas de error de Tipo I asociadas con el procedimiento de *WJ*, los resultados que aparecen recogidos en la Tabla 1 evidencian la honradez del enfoque de *WJ*, pues en cincuenta y nueve de las sesenta condiciones investigadas las tasas de error se encontraban dentro del criterio de ± 2SE. Apuntar si acaso, que cuando $n_{mín}/r(q-1)$ se fijó en 1.5 las tasas de error tendían a ser mayores.

En lo referido a la potencia de las pruebas, de los resultados esquematizados en la Tabla 1, conviene destacar, sobre manera, que ningún enfoque es uniformemente más potente que sus competidores. Sin embargo, por lo que respecta a los dos que arrojan resultados más parecidos, en concreto los de *BF* y *WJ*, el procedimiento de *WJ* resultó ser siempre más potente que el *BF*, excepción hecha de la condición en que la relación entre el tamaño de los grupos y el de las matrices de dispersión era negativa y el valor de $N = 54$. El enfoque *MDM*, que como se ha resaltado, no controlaba bajo ninguna de las condiciones examinadas las tasas de error de Tipo I, resultaba más potente que sus rivales cuando se comportaba de una manera liberal, esto es, cuando el diseño estaba equilibrado o cuando la relación entre el tamaño de los grupos y el tamaño de las matrices de dispersión era negativa, mientras que cuando dicha relación era positiva resultaba bastante menos potente que los otros enfoques considerados.

Tabla 2. Tasas empíricas de error y de potencia para la interacción grupos de tratamiento × ocasiones de medida.

				MDM		*BF*		*WJ*		*MDM*		*BF*		*WJ*	
Σ	n_1	n_2	n_3	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$	$\alpha = .05$	$\alpha = .01$
	12	12	12	0.0783	0.0181	0.0455	0.0091	0.1059	0.0292	0.5812	0.3180	0.4718	0.2095	0.5643	0.3044
	09	12	15	0.0309	0.0051	0.0499	0.0106	0.0942	0.0246	0.4287	0.1823	0.5555	0.2788	0.7415	0.4661
A	15	12	09	0.1512	0.0494	0.0246	0.0030	0.1476	0.0517	0.7633	0.5199	0.3663	0.1110	0.7180	0.4281
	18	18	18	0.0718	0.0197	0.0517	0.0118	0.0692	0.0152	0.8587	0.6476	0.8104	0.5576	0.9207	0.7432
	12	18	24	0.0079	0.0007	0.0496	0.0106	0.0651	0.0155	0.6416	0.3220	0.8659	0.6609	0.9172	0.7467
	24	18	12	0.1770	0.0627	0.0366	0.0067	0.0963	0.0271	0.9385	0.8176	0.6990	0.3871	0.8782	0.6482
	12	12	12	0.0708	0.0161	0.0463	0.0099	0.1049	0.0280	0.6234	0.3540	0.5206	0.2366	0.7348	0.4503
	09	12	15	0.0310	0.0076	0.0514	0.0110	0.0980	0.0271	0.4161	0.1832	0.5468	0.2718	0.7433	0.4705
B	15	12	09	0.1496	0.0515	0.0278	0.0031	0.1439	0.0498	0.7605	0.5166	0.3569	0.1168	0.7136	0.4290
	18	18	18	0.0719	0.0175	0.0532	0.0101	0.0709	0.0169	0.8560	0.6412	0.8113	0.5559	0.9237	0.7447
	12	18	24	0.0074	0.0010	0.0525	0.0109	0.0634	0.0154	0.6407	0.3257	0.8155	0.5620	0.9241	0.7452
	24	18	12	0.1826	0.0660	0.0396	0.0056	0.0943	0.0259	0.9384	0.8229	0.7006	0.3897	0.8771	0.6503
	12	12	12	0.0751	0.0194	0.0464	0.0085	0.1050	0.0274	0.6221	0.3602	0.5229	0.2415	0.7375	0.4567
	09	12	15	0.0316	0.0079	0.0490	0.0111	0.1012	0.0264	0.4351	0.1900	0.5585	0.2892	0.7440	0.4703
C	15	12	09	0.1526	0.0501	0.0257	0.0031	0.1503	0.0472	0.7644	0.5227	0.3702	0.1123	0.7179	0.4135
	18	18	18	0.0700	0.0184	0.0505	0.0114	0.0711	0.0137	0.8584	0.6517	0.8155	0.5620	0.9241	0.7490
	12	18	24	0.0094	0.0012	0.0512	0.0119	0.0664	0.0142	0.6438	0.3275	0.8734	0.6593	0.9190	0.7575
	24	18	12	0.1826	0.0662	0.0433	0.0067	0.0972	0.0283	0.9468	0.8253	0.7060	0.3897	0.8848	0.6545
	12	12	12	0.0728	0.0185	0.0452	0.0087	0.1020	0.0279	0.6223	0.3523	0.5160	0.2384	0.7381	0.4537
	09	12	15	0.0326	0.0043	0.0531	0.0096	0.0989	0.0263	0.4287	0.1815	0.5587	0.2778	0.7463	0.4760
D	15	12	09	0.1515	0.0500	0.0263	0.0036	0.1429	0.0478	0.7589	0.5112	0.3692	0.1127	0.7188	0.2444
	18	18	18	0.0650	0.0173	0.0481	0.0083	0.0693	0.0151	0.8622	0.6480	0.8134	0.5591	0.9227	0.7434
	12	18	24	0.0087	0.0027	0.0498	0.0110	0.0678	0.0143	0.6455	0.3264	0.8754	0.6600	0.9204	0.7470
	24	18	12	0.1854	0.0690	0.0423	0.0067	0.1030	0.0294	0.9467	0.8265	0.7014	0.3903	0.8838	0.6579
	12	12	12	0.0722	0.0182	0.0454	0.0091	0.1080	0.0289	0.6380	0.3708	0.5366	0.2588	0.7525	0.4761
	09	12	15	0.0321	0.0062	0.0518	0.0112	0.0966	0.0261	0.4573	0.2087	0.5841	0.3054	0.7643	0.4922
E	15	12	09	0.1659	0.0559	0.0285	0.0027	0.1492	0.0506	0.7797	0.5446	0.3860	0.1182	0.7258	0.4432
	18	18	18	0.0700	0.0175	0.0487	0.0103	0.0681	0.0131	0.8735	0.6779	0.8350	0.5890	0.9352	0.7725
	12	18	24	0.0090	0.0010	0.0502	0.0115	0.0632	0.0163	0.6965	0.3765	0.8995	0.7113	0.9371	0.7904
	24	18	12	0.1845	0.0662	0.0397	0.0061	0.0959	0.0311	0.9470	0.8413	0.7171	0.4151	0.8900	0.6703

MDM = Modelo doblemente multivariado; *BF* = Procedimiento de Brown-Forsythe; *WJ* = Procedimiento de Welch-James.

4.2. Tasas empíricas de error de Tipo I y de potencia para la interacción grupos $\times$ ocasiones

La Tabla 2 contiene las tasas empíricas de error y de potencia correspondientes a la interacción para los tres procedimientos analíticos, bajo cada una de las diferentes condiciones manipuladas. Por lo que se refiere a las tasas de error de Tipo I asociadas con el enfoque *MDM*, los resultados obtenidos para la interacción son muy similares a los presentados en la Tabla 1 para el efecto principal correspondiente a las ocasiones de medida bajo el mismo procedimiento; por tanto, evitaremos volver a hacer hincapié en lo ya dicho.

A su vez, los resultados de la Tabla 2 asociados con el enfoque *BF* revelan, que si bien dicho enfoque controla adecuadamente la tasa de error en la mayoría de las situaciones estudiadas, existen algunos casos donde esto no sucede. Como se puede apreciar en la Tabla 2, siempre que la relación entre el tamaño de los grupos y el tamaño de las matrices de dispersión era negativa el procedimiento de *BF* se comportaba de una manera conservadora; no obstante, dicha conducta no permanecía estable, ya que el grado de conservadurismo tendía a disminuir al incrementarse el tamaño de muestra. Por ejemplo, para $\alpha = 0.05$ el valor de $\hat{\alpha}$ promediado a lo largo de las distintas matrices de varianzas-covarianzas fue de 0.0265 cuando $n_1 = 15$, $n_2 = 12$ y $n_3 = 9$, mientras que para el mismo nivel de significación el valor de $\hat{\alpha}$ fue 0.0403 cuando $n_1 = 24$, $n_2 = 18$ y $n_3 = 12$. A pesar de lo dicho, en lo referido a la interacción, el enfoque *BF* exhibe un comportamiento muy superior al resto procedimientos evaluados en este trabajo.

Con respecto al comportamiento del enfoque *WJ*, es notorio que los datos recogidos en la Tabla 2 realzan la excesiva liberalidad exhibida por el enfoque *WJ* a lo largo de todas las condiciones analizadas, de manera especial, cuando la relación entre el tamaño de los grupos y el tamaño de las matrices de dispersión era negativa. La incapacidad del enfoque *WJ* para controlar las tasas de error cuando el interés se centra en los efectos de la interacción no deja de ser chocante, al menos, si se tiene presente su robustez para controlar las tasas de error de Tipo I referidas a las ocasiones de medida.

Por último, en lo que a la potencia de las pruebas se refiere, los resultados recogidos en la Tabla 2 ponen de relieve varios aspectos. En primer lugar, que los diferentes procedimientos examinados tenían una mayor sensibilidad para captar la interacción entre los grupos de tratamiento y las ocasiones de medida que para captar las diferencias entre estas últimas. En segundo lugar, que ningún enfoque resultó uniformemente más potente que sus competidores, aunque, en líneas generales, el procedimiento de *WJ* fue el que mostró una mayor sensibilidad estadística y el de *BF* el que menor. Sin embargo, como se resaltó en el párrafo anterior, no conviene perder de vista que bajo H_0 el enfoque *WJ*, al revés de lo que ocurría con el de *BF*, no fue robusto para ninguna de las condiciones manipuladas. En tercer lugar, el enfoque *MDM*, que como también se ha resaltado tampoco controlaba las tasas de error de Tipo I bajo ninguna de las condiciones examinadas, resultaba más potente que el de *BF* cuando el diseño estaba equilibrado o cuando la relación entre los tamaños de los grupos y el tamaño de las matrices de dispersión era negativa, mientras que cuando dicha relación era positiva la sensibilidad estadística era sustancialmente menor que la del procedimiento *BF*.

5. DISCUSIÓN Y CONCLUSIONES

El propósito de este trabajo fue examinar el comportamiento de los procedimientos *MDM*, el *WJ* y *BF* cuando se incumplía el supuesto de homogeneidad de las matrices de dispersión en un diseño multivariado de medidas parcialmente repetidas.

Con independencia de la fuente de variación implicada, los resultados de nuestro estudio indican la falta de robustez del enfoque *MDM* a la violación del supuesto de homogeneidad de las matrices de dispersión; en particular, cuando el diseño no estaba equilibrado. El descubrimiento de que el enfoque *MDM* no es viable cuando las matrices de dispersión difieren entre sí, no hace más que confirmar y generalizar los descubrimientos hallados por numerosos investigadores (Keselman y Keselman, 1990; Keselman y Lix, 1997; Vallejo *et al.* 1999), tanto cuando las medidas repetidas eran obtenidas en la perspectiva univariadada como multivariada. Al hilo de lo dicho, tenemos que concluir que el enfoque *MDM* no es robusto, lo cual contraviene la afirmación de Keselman y Lix (1997, p.333) de que, en muchos casos, el procedimiento controla adecuadamente las tasas de error. Pues bien, la razón de esta aparente discrepancia es más de forma que de fondo y no hay que buscarla en el distinto grado de heterogeneidad de las matrices (en ambos estudios la razón entre los elementos de las matrices fue de 1:3:5), sino más bien en los distintos tamaños de *N* utilizados en ambos estudios y, sobre todo, en la distinta medida utilizada para cuantificar la robustez; de haberse empleado el mismo criterio en ambos estudios las conclusiones serían muy parecidas. Cuando el diseño no es equilibrado, carece de sentido discutir si el enfoque controla o no la tasa de error escudándonos en utilizar un criterio liberal o conservador para medir la robustez. En esta situación el enfoque carece, claramente, de robustez y lo único que le cabe al investigador es decidir si opta por su abandono o por modificar los niveles de significación. Nosotros, a la luz de los resultados obtenidos recomendamos la primera solución, pues se nos antoja muy complicado sugerir una cantidad por la cual el investigador debería de multiplicar o dividir su nivel de significación nominal en función de la naturaleza de la relación existente entre el tamaño de los grupos y el tamaño de las matrices de dispersión.

Afortunadamente, a la hora de probar las hipótesis referidas a las fuentes de variación del diseño multivariado de medidas parcialmente repetidas, otras soluciones están disponibles (Wolfinger, 1996) y dos de ellas fueron aplicadas por nosotros. En lo tocante al efecto principal de las ocasiones de medida, los resultados ofrecidos en este trabajo indican que el procedimiento de *WJ* proporciona un excelente control de las tasas de error de Tipo I bajo las diferentes condiciones manipuladas. Así pues, aunque Keselman *et al.* (1993) recomienden que el enfoque no debería utilizarse cuando el tamaño de muestra sea pequeño, los datos de nuestro estudio ponen de relieve que, al menos en lo referido a los efectos principales, tal vez sólo sea necesario cumplir con el requisito matemático de que el tamaño de los grupos exceda al de variables dependientes multiplicado por el número de medidas repetidas menos uno; esto es, que n sea mayor que $r(q-1)$ cuando el diseño esté equilibrado y que $n_{mín}$ sea mayor que $r(q-1)$ cuando no lo esté. Por su parte el procedimiento *BF* también ejercía un aceptable control de las tasas de error cuando el diseño estaba equilibrado o la relación entre el tamaño de los grupos y el tamaño de las matrices de dispersión era negativo; sin embargo, el procedimiento *BF* tenía un comportamiento conservador cuando la relación entre el tamaño de los grupos y el tamaño de las matrices de dispersión era positiva. Además, el grado de conservadurismo en vez de disminuir al incrementar la razón $n_{mín}/r(q-1)$ de 1.5 a 2 se acrecentaba. Este último aspecto, unido a la mayor potencia mostrada por el enfoque *WJ*, nos lleva a que recomendamos utilizar el procedimiento *WJ* en orden a probar el efecto principal de las ocasiones de medida.

Finalmente, en lo que a los efectos de la interacción se refiere conviene destacar que, para condiciones similares a las manipuladas en nuestro estudio, el procedimiento *BF* proporciona un control de la tasa de error de Tipo I, sustancialmente mejor que el procedimiento *WJ*. Así pues, a la hora de probar hipótesis referidas a la interacción, nuestros resultados sugieren claramente que el procedimiento de *WJ* no debería ser utilizado, al menos, para tamaños de muestra similares a los empleados en este trabajo. Por consiguiente, dada la robustez del procedimiento *BF* a la falta de homogeneidad de las matrices de covarianza, todo indica que en estos casos el procedimiento a elegir debería ser el de *BF*. Con todo, en base a los recientes descubrimientos de Keselman y Lix (1997), conviene no olvidar que es posible obtener pruebas robustas del efecto de la interacción mediante el procedimiento *WJ* siempre y cuando la razón entre el tamaño de los grupos y el número de variables dependientes multiplicado por el de medidas repetidas menos una sea igual a 4 ó 5, dependiendo de si el diseño está o no equilibrado. En este caso, se plantea una nueva cuestión que, obviamente, dejaremos sin resolver, como es la referida a la sensibilidad estadística de las pruebas, pues cabe esperar que ambos procedimientos se comporten de una manera robusta cuando $n/r(q-1) \geqslant 4$ y $n_{mín}/r(q-1) \geqslant 5$. Otra cuestión sin resolver en el presente trabajo, y que sin duda puede afectar al alcance de las conclusiones actuales, tiene que ver con la estabilidad de los resultados obtenidos de utilizarse matrices de dispersión que no fuesen múltiplos unas de otras. No obstante, estamos seguros que futuras investigaciones se encargarán de esclarecerlo.

6. REFERENCIAS

Algina, J. (1994). «Some alternative approximate tests for a split-plot design». *Multivariate Behavioral Research*, 29, 365-384.

Boik, R.J. (1988). «The mixed model for multivariate repeated measures: Validity conditions and an approximate test». *Psychometrika*, 53, 469-486.

Boik, R.J. (1991). «Scheffé's mixed model for multivariate repeated measures: A relative efficiency evaluation». *Communication Statistics-Theory and Methods*, 20, 1233-1255.

Brown, M.B. and Forsythe, A.B. (1974). «The small sample behavior of some statistics which test the equality of several means». *Technometrics*, 16, 129-132.

Coombs, W.T. and Algina, J. (1992). *Four news solutions to the multivariate G-sample Behrens-Fisher problem*. Paper presented at the meeting of the Psychometric Society, Columbus, Ohio.

Coombs, W.T. and Algina, J. (1996). «New test statistics for MANOVA/descriptive discriminant analysis». *Educational and Psychological Measurement*, 58, 382-402.

Cornell, J.E., Young, D.M. and Bratcher, T.L. (1991). «An algorithm for generating covariance matrices with specified departures from sphericity». *Journal of Computation and Simulation*, 34, 240-243.

Davidson, M.L. (1972). «Univariate versus multivariate test in repeated measures experiments». *Psychological Bulletin*, 77, 446-452.

GAUSS (1997). *The Gauss System* (Vers. 3.2.32). Washington: Aptech Systems, Inc.

Greenhouse, S.W. and Geisser, S. (1959). «On methods in the analysis of profile data». *Psychometrika*, 24, 95-112

Johansen, S. (1980). «The Welch-James approximation of the distribution of the residual sum of squares in weighted linear regression». *Biometrika*, 67, 85-92.

Keselman, H.J., Carriere, M.C. and Lix, L.M. (1993). «Testing repeated measures hypotheses when covariance matrices are heterogeneous». *Journal of Educational Statistics*, 18, 305-319.

Keselman, J.C. and Keselman, H.J. (1990). «Analysing unbalanced repeated measures designs». *British Journal of Mathematical and Statistical Psychology*, 43, 265-282.

Keselman, H.J. and Lix, L.M. (1997). «Analysing multivariate repeated measures designs when covariance matrices are heterogeneous». *British Journal of Mathematical and Statistical Psychology*, 50, 319-338.

Kinderman, A.J. and Ramage, J.G. (1976). «Computer generation of normal random numbers». *Journal of the American Statistical Association*, 77, 893-896.

Nel, D.G. and van der Merwe, C.A. (1986). «A solution to the multivariate Behrens-Fisher problem». *Communications in Statistics-Theory an Methods*, 15, 3719-3735.

Rao, C.R. (1951). «An asymptotic expansion of the distribution of Wilks's criterion». *Bulletin of the International Statistical Institute*, 33, Part 2, 177-180.

Satterthwaite, F.E. (1941). «Synthesis of variance». *Psychometrika*, 6, 309-316.

Scheffé, H. (1956). «A mixed model for the analysis of variance». *Annals of Mathematical Statistics*, 27, 23-36.

Schauer, E.M. and Stoller, D.S. (1966). «On the generation of normal random vectors». *Technometrics*, 4, 278-280.

Tang K.L. and Algina, J. (1993). «Performance of four multivariate tests under variance covariance heteroscedasticity». *Multivariate Behavioral Research*, 28, 391-405.

Vallejo, G. y Escudero, J.R. (1998). «Algunas soluciones aproximadas para diseños split plot con matrices de dispersión arbitrarias». *Qüestiió: Quaderns d'Estadística i Investigació Operativa*, 22(3), 443-468.

Vallejo, G., Fernández, P., Fidalgo, A.M. y Escudero, J.R. (1999). «Comparación de la robustez de cuatro pruebas en un diseño multivariado split-plot». *Metodología de las Ciencias del Comportamiento*, 1, 1-23.

Vallejo, G., Fidalgo, A.M. y Fernández, P. (1998). «Efectos de la no esfericidad en el análisis de diseños multivariados de medidas repetidas». *Anales de Psicología*, 14, 249-268.

Vallejo, G. y Menéndez, I.A. (1997). «Una comparación de enfoques alternativos para el análisis de diseños multivariados de medidas repetidas». *Psicothema*, 9, 647-656.

Wilks, S.S. (1932). «Certain generalizations in the analysis of variance». *Biometrika*, 24, 471-494.

Wolfinger, R.D. (1996). «Heterogeneous variance-covariance structures for repeated measures». *Journal of Agricultural, Biological, and Environmental Statistics*, 1(2), 205-230.

ENGLISH SUMMARY

TESTING OF HYPOTHESES IN MULTIVARIATE SPLIT-PLOT DESIGNS WITH ARBITRARY DISPERSION MATRICES⋆

G. VALLEJO SECO
J. R. ESCUDERO GARCÍA
A. M. FIDALGO ALISTE
M. P. FERNÁNDEZ GARCÍA
Universidad de Oviedo*

The aim of this paper is to evaluate several multivariate procedures for testing omnibus null hypotheses for data gathered from multivariate split plot designs when the assumption of homogeneity of dispersion matrices is violated. An examination of these procedures for a wide number of variables confirms, on the one hand, the robustness of the Welch-James multivariate solution given by Johansen (1980) for testing the trials main effect hypothesis and, on the other hand, the robustness of the multivariate generalization of the Brown-Forsythe (1974) procedure for testing the groups x trial interaction. Our results also show that the power advantage was small in conditions in which the Welch-James and Brown-Forsythe tests control the Type I error rates.

Keywords: multivariate split-plot designs, arbitrary dispersion matrices, robust and powerful procedures

AMS Classification (MSC 2000): 62K10, 62J1

⋆ This work has been supported by DGCYT (PS95-0228).
* Departamento de Psicología. Universidad de Oviedo. Plaza Feijoo, s/n. 33003 Oviedo. gvallejo@sci.cpd.uniovi.es.

– Received March 1999.
– Accepted March 2000.

In analyzing data from repeated measures designs with multiple dependent variables, either a multivariate mixed model (MMM; the Scheffé's mixed model generalized for application to multivariate case) or doubly multivariate model (DMM) perspective may be used. Both analyses require multivariate normality and variance homogeneity. Besides that, Boik (1988) showed that, given the two previous conditions, a condition called multivariate sphericity of the covariance matrix is both necessary and sufficient for the validity of the MMM analysis. Simulation studies showed that the unadjusted *MMM* test cannot be recommended except when multivariate sphericity is known to hold (e.g., Boik, 1991; Vallejo & Menéndez, 1997). One situation in which the adjusted *MMM* test is more powerful than the *DMM* test occurs when sample size is very small; however, if sample size is large, the *DMM* test must be preferred (Boik, 1991; Vallejo & Menéndez, 1997; Vallejo, Fidalgo, & Fernández, 1998).

Vallejo, Fernández, Fidalgo, and Escudero (1999) evaluated the power and robustness for the *DMM* test and the ε_2-corrected *MMM* test suggested by Boik (1991) in the presence of heteroscedasticity of the variance-covariance matrices and when data were non-normal in form under null and non-null hypothesis. Their results revealed that these tests were extremely sensitive to departures from covariance homogeneity when the design was unbalanced and the sample size was small. When the design was balanced both adjusted degrees of freedom univariate or multivariate approaches were generally robust to covariance heterogeneity. Data distribution had small effects on the Type I error rates and power for both procedures: the *DMM* test was slightly liberal when the model was additive and conservative when the model was nonadditive; its effect for ε-corrected *MMM* tests was small.

Keselman, Carriere, and Lix (1993) used results in Johansen (1980) to develop a Welch-James (WJ) multivariate test for testing the main and interaction effects in unbalanced repeated measures designs without assuming covariance homogeneity. Results in Keselman and Keselman (1993) showed that, when the main effect is tested, the *WJ* test controls the Type I error rate. However, for testing the interaction effect, the WJ test does not control the Type I error rate. The problem was most extreme when the largest group size was associated with the covariance matrix containing the smallest element values. Keselman presented similar results and Lix (1997) did so for the multivariate repeated measures designs.

In summary, the literature indicates that the *WJ* test for assessing the repeated measures interaction effect cannot be recommended when the design is unbalanced and the covariance matrices are heterogeneous if sample size is small. In order to circumvent such problems, the multivariate Brown and Forsythe (BF; 1974) test of the interaction effect, proposed by Coombs and Algina (1992), is a promising alternative to the *WJ* test, according to the results obtained by Algina (1994). Accordingly, the major purposes of this study were to compare the Type I error and power rates of the *DMM*, *WJ* and modified *BF* statistics for testing the between $\times$ within interaction and the within-subjects main effect in multivariate repeated measures designs when covariance matrices were heterogeneous and the nonadditive model.

The study findings indicate that the *DMM* statistic cannot be recommended to test main and interaction effect hypothesis because it was not able to control the rate of Type I errors across all of the investigated conditions, especially when the design was unbalanced. These results are consistent with Keselman and Keselman's (1990), Keselman and Lix's (1997), and Vallejo et al's (1998) findings.

Fortunately, researchers have alternative analytic strategies: the *BF* and the *WJ* tests. The results indicate that, for testing the repeated measures main effect hypothesis, the W-J test can control the Type I error rate. As well, this control can be obtained even with small sample sizes whenever $n > r(q-1)$ for balanced design or $n_{mín} > r(q-1)$ for unbalanced designs. Likewise, the *BF* test was robust when the design was balanced or the pairing pattern was negative. However, the *BF* test was always conservative for positive pairings. Moreover, the degree of conservativeness increased when the ratio $n_{mín}/r(q-1)$ increased from 1.5 to 2. Because the *WJ* procedure was more robust and displayed significant power advantage over the *BF* test, we recommended that for testing the repeated measures main effect the *WJ* test be used.